Crimes in Philadelphia

Raj Patel, Hong Son, Kunal Sharma
Drexel University, 2020-2021 Fall Quarter
DSCI591 Data Science Capstone I

Identifying a Dataset

Philadelphia Crime Data:

https://www.kaggle.com/mchirico/philadelphiacrimedata

For our project, we would like to conduct an analysis on a Philadelphia crime dataset published by OpenPhillyData. The dataset dates from 12-31-05 to 03-22-17 (majority of the dates are from 2006-2016). There are 14 total columns in this dataset. They include the district number, sector, dispatch date, dispatch time, location, general code, descriptions of the crime, and the district. This dataset was published to the public so people can assess the crimes happening in Philadelphia. For our exploratory data analysis, we will perform data cleansing, use data transformation to create new variables, and then create visualizations to answer the questions posed in this document

Weather Data:

http://www.climate.psu.edu/data/city_information/index.php?city=phl&page=dwa&type=big7

To assist in getting more out of our analysis, we will be scraping Philadelphia weather data for Philadelphia from 2007 to 2018. The dataset includes the temperatures in Fahrenheit, the max and min temperature, environment information, wind, and precipitation. The data was formatted into PDF files and a dataset was formed [~4,000 rows and 15 columns].

Understanding the Target Data and Substantiating its Existence

Potential users of this analysis are law enforcement agencies so that they can dedicate resources and assistance to areas with high severity of crimes. They will also be able to assess the crime rates and find trends in crimes. Our proposed solution is to create a model (If applicable) that can predict the crimes in order to help the agencies plan out their resources in order to lessen the number of crimes and make Philly safer.

We will attempt to answer the following questions in our analysis:

1. What are the top 10 crimes over the 10 year period ?
2. Have crimes been increasing over the 10 year period?
3. What types of crimes are on the rise? What types of crimes have fallen? (Over 10 year period)
4. Create a heat map with the severity of the crimes (ie. thefts) to answer questions around the areas with the highest counts of the associated crimes.
5. What District/Region has the most crimes?
6. Does the season affect the number of crimes?
7. What are the dangerous times (hours) to be outside or for crimes to happen?
8. If applicable, predict the number of crimes related to the top 3 incidents.
9. Given a location, what kind of crime occurs the most?
10. Is there a pattern to the occurrence of crime? Example: a day of a week when it is more likely that a crime will occur, the average number of crimes likely to occur in a day, week or month.

INTENDED STEPS
- Finish & Revise Outline Document
- Weekly Team meeting to Discuss Project
- Find and retrieve all relevant datasets required in the project
- Process the Data for Analysis - Cleansing
    - Missing Values
    - Corrupted Values
    - Errors
- *Data Acquisition and Preprocessing Report*
- Explore the data (Data Combining, Statistical Analysis)
- Data Visualizations
- Perform in-depth analysis
- Communicate Results
- *Exploratory Data Analytics Report Draft*
- Final Project Presentation